Complete genome sequence of the prototype lactic acid bacterium Lactococcus lactis subsp cremoris MG1363

Wegmann, Udo; O'Connell-Motherwy, Mary; Zomer, Aldert; Buist, Girbe; Shearman, Claire; Canchaya, Carlos; Ventura, Marco; Goesmann, Alexander; Gasson, Michael J.; Kuipers, Oscar P.

Published in:
Journal of Bacteriology

DOI:
10.1128/JB.01768-06

Publication date:
2007

JOURNAL OF BACTERIOLOGY, Apr. 2007, p. 3256–3270 Vol. 189, No. 8
0021-9193/07/$08.00+0 doi:10.1128/JB.01768-06

Complete Genome Sequence of the Prototype Lactic Acid Bacterium *Lactococcus lactis* subsp. *cremoris* MG1363▽

Udo Wegmann,[1]§ Mary O'Connell-Motherway,[2]§ Aldert Zomer,[3]†§ Girbe Buist,[3]‡ Claire Shearman,[1] Carlos Canchaya,[2] Marco Ventura,[2] Alexander Goesmann,[4] Michael J. Gasson,[1]* Oscar P. Kuipers,[3]* Douwe van Sinderen,[2] and Jan Kok[3]*

Institute of Food Research, Norwich Research Park, Colney, Norwich, NR4 7UA, United Kingdom[1]; Alimentary Pharmabiotic Centre and Department of Microbiology, National University of Ireland, Cork, Ireland[2]; Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Haren, 9751 NN, The Netherlands[3]; and Center for Biotechnology, Bioinformatics Resource Facility, Universität Bielefeld, 33594 Bielefeld, Germany[4]

Received 21 November 2006/Accepted 6 February 2007

***Lactococcus lactis* is of great importance for the nutrition of hundreds of millions of people worldwide. This paper describes the genome sequence of *Lactococcus lactis* subsp. *cremoris* MG1363, the lactococcal strain most intensively studied throughout the world. The 2,529,478-bp genome contains 81 pseudogenes and encodes 2,436 proteins. Of the 530 unique proteins, 47 belong to the COG (clusters of orthologous groups) functional category "carbohydrate metabolism and transport," by far the largest category of novel proteins in comparison with *L. lactis* subsp. *lactis* IL1403. Nearly one-fifth of the 71 insertion elements are concentrated in a specific 56-kb region. This integration hot-spot region carries genes that are typically associated with lactococcal plasmids and a repeat sequence specifically found on plasmids and in the "lateral gene transfer hot spot" in the genome of *Streptococcus thermophilus*. Although the parent of *L. lactis* MG1363 was used to demonstrate lysogeny in *Lactococcus*, *L. lactis* MG1363 carries four remnant/satellite phages and two apparently complete prophages. The availability of the *L. lactis* MG1363 genome sequence will reinforce its status as the prototype among lactic acid bacteria through facilitation of further applied and fundamental research.**

Lactococcus lactis, a mesophilic fermentative bacterium producing lactic acid from sugar (hexose) fermentation, is an important industrial microorganism with extensive and diverse uses in food fermentation. Strains of *L. lactis* are used as defined mixtures or in undefined combinations with other lactic acid bacteria (LAB) in the production of fermented milk products. The organism has adapted to growth in milk under stringent human selection for better performance with respect to taste, flavor, and texture of dairy products, and this process continues today (57, 98, 99). In 1985, the "dairy streptococci" were reclassified into two *L. lactis* subspecies, *Lactococcus lactis* subsp. *lactis* (previously *Streptococcus lactis*) and *Lactococcus lactis* subsp. *cremoris* (previously *Streptococcus cremoris*), to distinguish them from the streptococci sensu stricto, which contain a number of notorious human pathogens (82, 83).

The strain used in this study, *L. lactis* subsp. *cremoris* MG1363, is the international prototype for LAB genetics, and the knowledge gained from fundamental research on this strain has been exploited for a wide variety of biotechnological applications. The large and unstable complement of plasmid DNA of the parent strain, *L. lactis* NCDO712, was eliminated by employing UV treatment and protoplast-curing strategies in the early 1980s (41). The resultant plasmid-free strain, *L. lactis* MG1363, is robust and genetically amenable, which has facilitated the analysis of introduced lactococcal and heterologous DNA. Sophisticated systems have been developed for the expression of proteins and peptides in this strain, and it has been used as a cell factory for a wide variety of heterologous products (e.g., antimicrobials, including bacteriocins [50], bacteriophage endolysins [75], and defensins [47]). A recent review provides a snapshot of this diversity involving the favored "NICE" system (for nisin-controlled protein overexpression) (25, 71), which can be used to express heterologous proteins (e.g., lysostaphin) up to industrial scale (72). The importance of these developments for the field of biotechnology and for microbiology research is enormous, as many of the tools initially developed for *L. lactis*, e.g., plasmids, integration systems, and the NICE expression system, have also been shown to be applicable in all other LAB (55) and several species of other industrially relevant bacterial genera, e.g., *Bacillus* and *Clostridium*, and also in human pathogens, e.g., *Listeria*, *Enterococcus*, and *Streptococcus* (14, 58, 79).

The GRAS (generally regarded as safe) status of *L. lactis* is a distinct advantage for its use in the production and secretion of therapeutic or vaccine proteins (62). While *L. lactis* is not a natural inhabitant of the gastrointestinal tract, it does survive gut passage. *L. lactis* MG1363 has been successfully used to pioneer the gut delivery of bioactive molecules, such as vaccine antigens and immune modulators. It is especially relevant that a contained mutant of the genetically modified strain expressing the cytokine interleukin 10 has been used in a human trial of patients with inflammatory bowel disease (88–90). A muco-

* Corresponding author. Mailing address: Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands. Phone: 31 50 363 2111. Fax: 31 50 363 2348. E-mail: jan.kok@rug.nl.

† Present address: Alimentary Pharmabiotic Centre and Department of Microbiology, National University of Ireland, Cork, Ireland.

‡ Present address: Laboratory of Molecular Bacteriology, Department of Medical Microbiology, University Medical Center Groningen (UMCG), Groningen, The Netherlands.

§ U.W., M.O.-M., and A.Z. contributed equally to this work.

▽ Published ahead of print on 16 February 2007.

sal vaccine based on live *L. lactis* MG1363 expressing the E7 antigen and interleukin 12 was shown to protect mice against human papillomavirus type 16-induced tumors (6). Recently, *L. lactis* was used as a nonliving nonrecombinant delivery system for mucosal vaccination (3). Pneumococcal antigens were bound to the walls of pretreated lactococcal cells by means of the peptidoglycan-binding domain of the major lactococcal autolysin (87). Local and systemic immune responses were induced following intranasal immunization of mice.

The analysis and engineering of lactococcal metabolism has been focused on *L. lactis* MG1363 and has involved both homologous pathways and the introduction of heterologous genes to reroute metabolic flux (26, 85). The relative simplicity of the lactococcal metabolism lends itself to modeling, e.g., derivatives of the strain that overproduce B vitamins (92) or alanine (49) have been constructed through metabolic rerouting.

In summary, in the international scientific community, *L. lactis* MG1363 has deserved its status as the undisputed prototype LAB because of the extent and diversity of its combined genetic and biotechnological amenities. The *L. lactis* MG1363 genome sequence presented here is vital to both applied and fundamental research, allowing the application of powerful postgenomic techniques (e.g., proteomics, transcriptomics, and metabolomics) to further these studies.

MATERIALS AND METHODS

Genome sequencing and assembly. Genomic DNA from *L. lactis* subsp. *cremoris* MG1363 was used to construct a small insert library (1 to 3 kb) in the pGEM-T (Promega, Madison, WI) vector. DNA sequencing was performed using standard primers on automated sequencing machines (ABI373 and ABI3730 [Applied Biosystems, Foster City, CA] and Licor [LI-COR Biosciences]). Base calling was performed with the phred package (38), followed by assembly using the phrap package (http://www.phrap.org) in conjunction with the Staden package (86). The resulting draft assembly was then mapped on the genome of *L. lactis* IL1403 using the *Projector* software package (97) to order the contigs into large scaffolds. Standard PCR, followed by primer walk sequencing of the resulting products, was used to close the remaining 800 gaps, in combination with inverse PCR for the gaps for which no linkage information could be obtained. To verify the complete assembly, PCRs predicted to be separated by 9 kb were performed on the genome, covering the entire genome with an overlap of 500 bp. Virtual digests of the complete assembly were compared with available pulsed-field gel electrophoresis restriction maps of the genome of *L. lactis* MG1363 (61). All low-quality regions on the genome were resequenced until every base pair was at least of phred 30 quality, resulting in a final average quality of approximately phred 70 and fivefold coverage of the entire genome. Regions containing putative frameshifts and point mutations were resequenced to verify the fidelity of the sequence.

Bioinformatics analyses. The finished *L. lactis* MG1363 sequence was annotated using the GenDB 2.2 Annotation Tool (70). Putative protein coding sequences (CDSs) were determined using REGANOR5 (68), based on the combined CDS predictions of CRITICA (4) and Glimmer (80). Putative ribosomal binding sites and tRNA genes were identified with RBSfinder (91) and tRNAscan-SE (65), respectively. An automatic functional annotation was followed by a manual annotation of each predicted gene. Pseudogenes were identified through FastA analysis of the intergenic regions versus the nonredundant nucleotide sequence database and a comparison of predicted gene products with respective proteins in the nonredundant peptide sequence database. Similarity searches were performed using BLASTN and BLASTP (2) against the nonredundant nucleotide and protein databases, respectively, provided by the National Center for Biotechnology Information. Additionally, a BLASTP search was performed against the KEGG database (54), and PSI-BLAST (2) was used to perform similarity searches against the SWISS-PROT (8) database. Analysis of protein families was done using the HMMer package (36) for searches against the Pfam (5) and TIGRFAM (46) databases. Furthermore, functional motifs and protein domains were automatically assigned by InterProScan queries against InterPro (73). RPSBlast was used to search the conserved domain database CDD (67). Searches against TransTerm (13) were applied on intergenic regions to predict transcription terminators. Additionally, SignalP (35), helix-turn-helix (29), and TMHMM (59) were applied. Finally, each gene was functionally classified by assigning a "clusters of orthologous groups" (COG) number and corresponding COG category (94) and gene ontology numbers (48) based on the best BLASTP results versus COG and InterPro results, respectively. Similarity searches using BLASTP were performed against the toxin and virulence factor database MvirDB to assess the GRAS status of *L. lactis* MG1363 (103).

Predicted CDSs were manually reviewed, and alterations were made on the basis of the presence of potential ribosomal binding sites, sequence alignments, and available data in the literature. Potential alien genes on the *L. lactis* genome sequence were identified using the program SIGI (for score-based identification of genomic islands) (69), based on the scoring of codon frequencies combined with cluster analysis. Alien genes are defined as those that are potentially acquired through horizontal gene transfer (HGT). This program has also been used to analyze the genome sequence of *Bacillus licheniformis* DSM13 (101) and the spread of the *ycdB* gene from *Lactococcus* to *Salmonella* (10). In order to identify secreted proteins, all putatively expressed proteins of *L. lactis* MG1363 were analyzed using the SignalP algorithm, and potential transmembrane domains were determined using the TMHMM (v2.0) algorithm. Proteins containing a signal peptide and a single transmembrane domain were further screened for the presence of the following protein retention signals: a lipobox (for lipid attachment to the membrane), a prepilin motif (for the formation of [pseudo]pili, which would indicate that the protein remained anchored to the cell envelope), peptidoglycan- or choline-binding domains for noncovalent binding to the cell wall, and an LPXTG motif for covalent attachment to the cell wall mediated by sortase.

Carbon utilization. *L. lactis* MG1363 and *L. lactis* IL1403 were tested on Biolog Phenotype MicroArray plates for carbon catabolism (7).

Nucleotide sequence accession number. The whole genome sequence of *L. lactis* MG1363 (http://www.cebitec.uni-bielefeld.de/groups/brf/cooperations/LlactisMG1363.html) has been deposited in the EMBL/GenBank databases with the accession number AM406671.

TABLE 1. General genome features

Feature	Value in *L. lactis* strain:		
	MG1363	IL1403	SK11
Size (bp) of chromosome	2,529,478	2,365,589	2,438,589
GC %	35.8	35.4	36.8
No. of predicted CDSs	2517	2310	2658[a]
No. of IS elements	71	43	130[a]
No. of phage genes	178	221	193
Phage DNA size (kb)	134	175	132

[a] Including plasmid-encoded genes and IS elements.

RESULTS AND DISCUSSION

General genome features. The circular chromosome of *L. lactis* MG1363 contains 2,529,478 bases (2.53 Mb), with an average GC content of 35.8% (Table 1). The Oriloc software (39) was used for the identification of the putative *oriC* and *ter* regions. The *dnaA* gene located in the putative *oriC* region is preceded by an AT-rich region (71%) containing seven DnaA boxes, four on the forward strand and three on the reverse strand. Therefore, the first base pair of *dnaA* was designated position 1 on the circular chromosome of *L. lactis* MG1363. The putative terminus is at position 1255251 (Fig. 1). A total of 2,517 predicted CDSs were identified, the majority (79%) of which were on the leading strand of chromosome replication. For 1,574 (62%) of the deduced CDS products, a general or specific function was predicted. A comparison with the genome sequences of *L. lactis* IL1403 (11) and *L. lactis* subsp. *cremoris* SK11 (66) revealed that there is approximately 85% DNA sequence identity between the CDSs present in both *L. lactis*

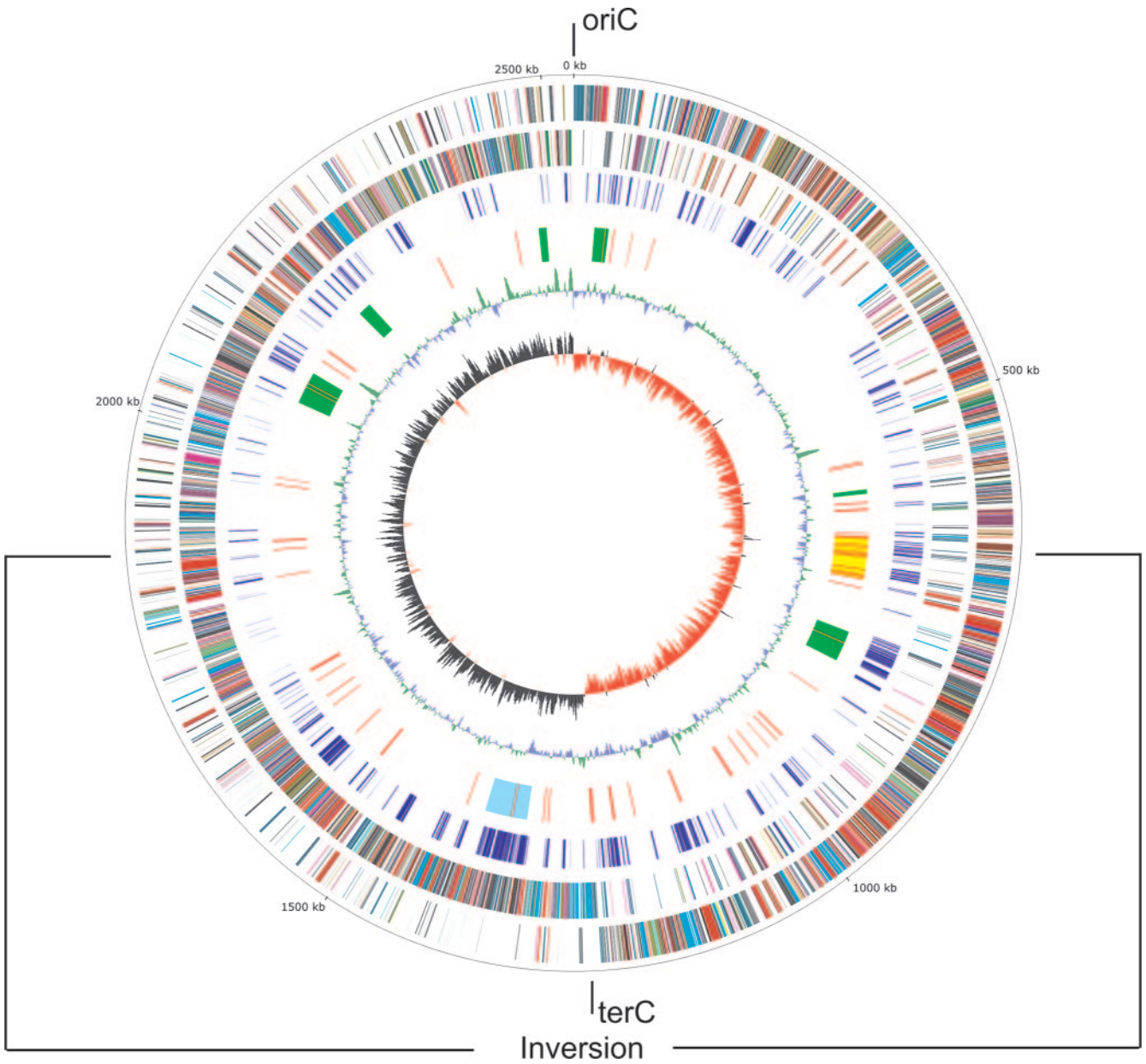

FIG. 1. Genome atlas of the chromosome of *L. lactis* MG1363. The first base pair of *dnaA* in the *oriC* region is designated position 1, and the putative terminus is at position 1255251. The inversion relative to *L. lactis* IL1403 is indicated. Starting from the inside, the circles depict the following. Circle no. 1 (red) represents the GC skew; the origin and terminus are clearly marked by the change in GC skew. Circle no. 2 (blue/green) shows the GC percentage of the CDSs on the genome, where green peaks are (six) tRNA genes or rRNA regions; regions with a lower GC content possibly originate from a lateral gene transfer event. Circle no. 3 marks the presence of IS elements (red) and prophages/phage remnants (green) and the location of the integration hot spot (yellow) and sex factor (light blue). Regions with many IS elements often coincide with sequences of relatively high AT, and IS elements are overrepresented in the integration hot spot. Circle no. 4 depicts the genes that are unique to *L. lactis* MG1363, i.e., that do not occur in *L. lactis* IL1403. The outer two circles are color coded according to the COG classification of the genes present on the forward (circle no. 6) and reverse (circle no. 5) strands.

MG1363 and *L. lactis* IL1403 and 97.7% between those in the two subspecies *cremoris* strains, *L. lactis* MG1363 and *L. lactis* SK11. The 16S rRNA sequences of *L. lactis* MG1363 and *L. lactis* IL1403 diverge by 0.33%, suggesting that they deviated approximately 17 million years ago (10). Using BLASTN at a cutoff of 1e−20, 465 of the *L. lactis* MG1363 genes do not have a corresponding hit in *L. lactis* IL1403 and 346 are not present in the chromosome of *L. lactis* SK11 (but 19 of these are present on plasmids in the latter), while 213 are absent in both *L. lactis* IL1403 and *L. lactis* SK11.

Eighty-one pseudogenes were detected in *L. lactis* MG1363 (Table 2), representing around 3% of the total number of genes, which is well within the 1% to 5% described for other bacterial genomes (64). Most of the pseudogenes (a total of 32) originate from genes involved in transposition, followed by genes with putative transport functions (a total of 11) and

TABLE 2. Pseudogenes in *L. lactis* MG1363 and putative functions of their unimpaired counterparts[a]

Locus tag	Gene	Putative function
llmg_pseudo_02		Putative transport protein
llmg_pseudo_03		Bacterial regulatory factor, effector
llmg_pseudo_04		HTH-type transcriptional regulator
llmg_pseudo_05	*dhaL*	Dihydroxyacetone kinase
llmg_pseudo_06	*glpF1*	Glycerol uptake facilitator
llmg_pseudo_07		ABC transporter ATP binding protein
llmg_pseudo_08	*enoB*	Enolase
llmg_pseudo_09	*dppP*	Dipeptide-binding protein
llmg_pseudo_10		Putative transport protein
llmg_pseudo_11	*kupA*	Putative potassium transport system protein
llmg_pseudo_14	*enoA*	Enolase
llmg_pseudo_17	*repB*	Replication protein
llmg_pseudo_23	*kdgK*	2-Keto-3-deoxygluconate kinase
llmg_pseudo_29		Putative transport protein
llmg_pseudo_30	*dxsA*	Deoxyxylulose-5-phosphate synthase
llmg_pseudo_32	*bglR*	Transcriptional antiterminator
llmg_pseudo_34	*bglR*	Transcriptional antiterminator
llmg_pseudo_37	*mnaA*	UDP-GlcNAc 2-epimerase
llmg_pseudo_39		HTH-type transcriptional regulator LacI family
llmg_pseudo_41	*ceo-2*	N5-carboxyethyl-ornithine synthase
llmg_pseudo_42	*leuB*	Isocitrate/isopropylmalate dehydrogenase
llmg_pseudo_43	*leuA*	2-Isopropylmalate synthase
llmg_pseudo_44	*tag*	DNA-3-methyladenine glycosidase
llmg_pseudo_46	*glf-2*	UDP-galactopyranose mutase
llmg_pseudo_47		Putative transport protein
llmg_pseudo_49	*glxK*	Glycerate kinase
llmg_pseudo_52		Putative sensor histidine kinase
llmg_pseudo_53		Putative response regulator
llmg_pseudo_54		Phosphotransferase system sugar-specific EIIC component
llmg_pseudo_57	*coiA*	Competence protein from the CoiA-like family
llmg_pseudo_58		ABC-type peptide transport system, permease component
llmg_pseudo_59		Putative phosphatase
llmg_pseudo_60	*comEC*	Competence protein
llmg_pseudo_61	*fadA*	Acetyl-coenzyme A acetyltransferase
llmg_pseudo_64	*oppF2*	Oligopeptide transport ATP-binding protein
llmg_pseudo_65	*oppD2*	Oligopeptide transport ATP-binding protein
llmg_pseudo_69		Putative oxidoreductase
llmg_pseudo_71		Putative amidase
llmg_pseudo_74	*noxD*	NADH oxidase
llmg_pseudo_78	*malQ*	4-Alpha-glucanotransferase
llmg_pseudo_79	*malQ*	4-Alpha-glucanotransferase
llmg_pseudo_80	*patB*	Aminotransferase
llmg_pseudo_81	*patB*	Aminotransferase

[a] Genes related to transposition or with unknown functions are not listed.

regulatory roles (a total of 6). If transposon-related genes are ignored, most of the pseudogenes (a total of 31) arose from the inactivation of single-copy genes. The decay of genes following possible duplication events contributed 14 pseudogenes. Although *L. lactis* MG1363 hosts a large number of insertion sequence (IS) elements (see below and Table 3), only three pseudogenes (*bglR*, *fadA*, and *malQ*) are the result of IS insertions.

IS elements, HGT, and alien genes. *L. lactis* MG1363 carries 11 different IS elements involving a total of 67 kb of DNA (Table 3). Several of these elements are contained within sequences not present in *L. lactis* IL1403 (Fig. 1). A rather remarkable feature is apparent in their distribution: nearly one-fifth of the IS elements are concentrated in a specific region of 59 kb, while three novel IS species are present only in this part of the chromosome. This "integration hot spot" also carries possible plasmid inserts (see below). The presence of the four IS elements, previously described only on plasmids,

TABLE 3. IS elements of *L. lactis* MG1363, *L. lactis* IL1403, and *L. lactis* SK11

IS element	No. in *L. lactis* strain[a]:		
	MG1363	IL1403	SK11
IS*904*	9 (1)	9	7
IS*1077*	9 (9)	7	10 (1); 1[b]
IS*905*	14 (8)	1	13
IS*981*	16 (1)	10	30 (1); 2[b]
IS*982*	2 (1)	1	55 (1); 4[b]
IS*983*	0	15	0
IS*712*	8 (1)	0	0
IS-*LL6*	9 (9)	0	3[b]
IS*946*	1	0	2; 1[b]
IS*1216*	1	0	3; 2[b]
IS*1297*	1	0	7[b]
IS*1675*	1	0	0
Total	71	43	130

[a] Number in parentheses indicates pseudogenes.
[b] Number of plasmid-located copies of the indicated IS elements.

and plasmid-derived genes in the integration hot spot show that there has been a high level of gene exchange with the plasmid population of the parental strain in the past. Two ISs are unique to *L. lactis* MG1363: IS*712* and IS*1675* are not present in *L. lactis* IL1403 and *L. lactis* SK11. A striking feature in the last strain is that two-thirds of its 130 IS elements are IS*981* (30 copies) and IS*982* (55 copies).

Many of the IS elements in *L. lactis* MG1363 are now inactive, as shown by the fact that 42% of the pseudogenes were originally involved in transposition. Despite the fluidity of IS movement within the genome, there are very few (three, as mentioned above) direct effects on gene integrity. Insertion of IS elements can also result in the activation of downstream genes. Examples in *L. lactis* are those identified after a spontaneous IS insertion event, such as constitutive nisin production (27, 28), and those resulting from IS insertion under a strong selection pressure, such as the IS activation of the alternative *ldhB* gene in *ldhA* knockout mutant strains, enabling them to grow much better. More than one IS element has been shown to be capable of this activation, as independent strains have been isolated (IS*981* [12] and IS*905* [our unpublished results]).

The sex factor is a unique mobile genetic element present on the *L. lactis* MG1363 chromosome (Fig. 1) (reviewed in reference 42). It can conjugate into *L. lactis* IL1403, due to the presence of a potential *attB* site between *guaC* and *xpt* (positions 1159525 to 1159548 in the *L. lactis* IL1403 genome) (reference 42 and our unpublished results). The genes surrounding the putative *attB* site in *L. lactis* IL1403 are the same as those flanking the sex factor in *L. lactis* MG1363 (i.e., those surrounding the *attB* site in *L. lactis* MG1363, if the sex factor was lost spontaneously or temporarily excised). The *L. lactis* IL1403 genes are colinear in this region with those of *L. lactis* MG1363. There is no indication of the presence of a sex factor-related element at this position in *L. lactis* IL1403, nor does *L. lactis* SK11 contain a sex factor.

Although natural competence in *L. lactis* MG1363 has not been described, the organism is capable of acquiring novel DNA through conjugation or sex factor-mediated transfer or via bacteriophage transduction (reviewed in references 40, 42, and 43). Analyses of codon or amino acid sequences, as well as phylogenetic analyses, have been used to predict chromosomal regions obtained through HGT. In *L. lactis* MG1363, 7.1% (174/2,576) of the genes were defined as alien, i.e., acquired through HGT (using the SIGI software based on codon usage and cluster analysis), which compares well with the 10% obtained for *Lactobacillus plantarum* WCFS (56), defined as acquired by HGT using base composition analysis, but slightly lower than that predicted by SIGI for *Streptococcus pneumoniae* TIGR4 (13.0%), both *Streptococcus agalactiae* and *Streptococcus pyogenes* (10.9%), *Listeria innocua* (10.6%), *Listeria monocytogenes* (9.0%), and, in particular, *L. lactis* IL1403 (8.9%) (69).

Lactococcus is viable in a number of diverse environments, is plant associated, and is found in the gastrointestinal tracts of animals, insects, and humans, and also in fermented foods and feeds of dairy, meat, and plant origin, bringing it into close proximity to a wide variety of other microorganisms with a large reservoir of genes for potential gene transfer (63). There is evidence that gene transfer can occur within the intestinal tract of mice (51). The *ycdB* gene has an unknown function and is widespread in bacteria (some enteric bacteria have two copies) and some eukaryotes. Analysis suggested that the gene was transferred from *L. lactis* to *Salmonella* (37), and more recent research has shown that this region of DNA has been transferred between IL-like and MG-like lactococcal strains, possibly by a conjugal process (10).

L. lactis MG1363 has been used extensively as a model for future use of LAB in medical applications, e.g., as a (live) oral vaccine (6, 3, 87). These developments are fueled by the fact that *L. lactis* is innocuous and has an extensive record as a GRAS organism. From the genome sequence, there is evidence for past and potential for future gene transfer in *L. lactis* MG1363. In a preliminary search, we examined the presence in *L. lactis* MG1363 of proteins with homology to known virulence factors, using the MvirDB database (103). A number of hits against MVirDB were detected at a cutoff of 1e−20, but apart from genes that specify proteins involved in general metabolic pathways, only *fbpA*, encoding a putative fibronectin-binding protein, and a phage-encoded putative extracellular endonuclease (ps111) have been implicated in virulence in other organisms. FbpA might, by analogy to the activity of the protein in *Streptococcus gordonii*, be involved in adhesion to cell surfaces (18). On the other hand, fibronectin-binding proteins may also be important for probiotic action and might not be virulence factors per se (1). The absence of obvious virulence genes, such as those encoding hemolysins or toxins, corroborates the GRAS status of *L. lactis* MG1363. Clearly, the availability of the genome sequence will add another layer of safety to future studies using *L. lactis* MG1363 for medical purposes.

Prophages and prophage remnants. The *L. lactis* MG1363 chromosome harbors six regions that represent bacteriophage-related sequences (Fig. 1). Two sites appear to contain complete prophage genomes, designated phiT712 (42,085 bp) and MG-3 (44,200 bp). The remaining bacteriophage sequences, designated MG-1 (19,053 bp), MG-2 (6,019 bp), MG-4 (18,029 bp), and MG-5 (10,598 bp), appear to represent remnant or satellite phages. Together, the bacteriophage sequences encompass approximately 5.5% of the *L. lactis* MG1363 genome, representing a large portion of the observed genomic differences between *L. lactis* MG1363, *L. lactis* SK11, and *L. lactis* IL1403. The latter finding is a clear indication that lysogenic bacteriophages contribute significantly to genome variability within this species. The six prophage sequences occupy various positions on the *L. lactis* MG1363 genome, and their insertion did not appear to have resulted in gene interruptions. Comparative analyses between the six prophages present in *L. lactis* IL1403 and those found in *L. lactis* MG1363 revealed only one common integration site, i.e., that of MG-1 and bIL310, two phages that also display the highest level of homology and synteny. A similar phage is not present in *L. lactis* SK11 (66). Interestingly, the integration site of phage MG-3 is the same as that of the *L. lactis* SK11 phage in the intergenic region between CDSs LACR_2145 and LACR_2146. However, there is sparse and low similarity between the two phage sequences, suggesting the presence of an insertion site that can be the target of different phages.

It is noteworthy that two of the *L. lactis* MG1363 phages (phiT712 and MG-4) contain sequences that are homologous

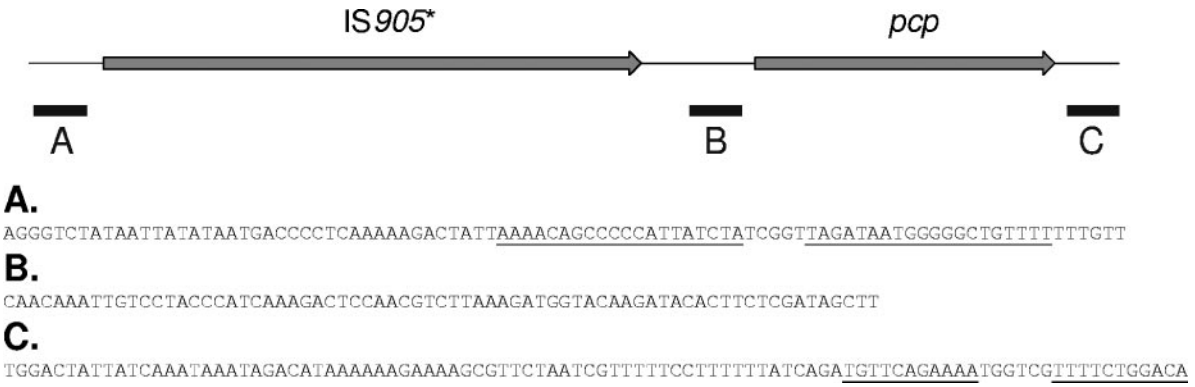

FIG. 2. Integration hot-spot repeat sequences. The underlined sequences indicate inverted repeats. The asterisk indicates the fact that the transposase gene of IS*905* is a pseudogene. (A) Sequence present in at least 27 plasmids from *L. lactis* subsp. *lactis*, *L. lactis* subsp. *cremoris*, and *L. lactis* bv. diacetylactis and plasmids of *Lactobacillus helveticus*, *Lactobacillus delbrueckii* subsp. *lactis*, *Tetragenococcus halophila*, *Pediococcus acidilactici*, *Leuconostoc citreum*, *L. monocytogenes*, and *L. innocua*, and also the genomes of *Lactobacillus sakei*, *Lactobacillus delbrueckii* subsp. *lactis*, and *Lactobacillus delbrueckii* subsp. *paracasei* and different strains of *S. thermophilus.* (B) Sequence present in at least nine plasmids from *L. lactis* subsp. *lactis*, *L. lactis* subsp. *cremoris*, and *L. lactis* bv. diacetylactis and plasmids of *L. helveticus*, *Lactobacillus lindneri*, *Lactobacillus brevis*, *L. sakei*, and *L. plantarum*, and also the genomes of *Lactobacillus paracasei* subsp. *paracasei* and *S. thermophilus.* (C) Sequence, or part thereof, present in at least 24 plasmids from *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris*. The inverted repeat is present in the genomes of different strains of *S. thermophilus* and *Acetobacter pasteurianus* and plasmids of *Staphylococcus epidermidis* (two), *S. aureus* (six), *Staphylococcus xylosus*, *Staphylococcus haemolyticus* (two), and *L. monocytogenes* and *L. innocua*.

to different plasmid-encoded abortive infection systems (reviewed in reference 17), which, together with a gene encoding a putative extracellular endonuclease found in prophage MG-1, seem to endow the host with a beneficial lysogenic conversion property. Moreover, prophages MG-2 and MG-3 carry tRNA genes (i.e., $tRNA^{Lys}$ and $tRNA^{Arg}$), which may increase the translational efficiency of the bacterial host. All *L. lactis* MG1363 prophages display limited similarity with DNA prophage sequences contained in *L. lactis* SK11. Their sequence similarity is mainly restricted to presumed integrase genes, to genes encoding structural proteins (e.g., phiT712), or to a putative DNA replication region (e.g., MG-4 and MG-3).

L. lactis MG1363 is a "phage-cured" strain; induction of prophages in *L. lactis* MG1363 has never been reported in the literature. It is worth noting that the largest prophage genomes in *L. lactis* MG1363 (MG-1, MG-3, and phiT712) contain IS*712* elements, which are unique to *L. lactis* MG1363. It is not known whether their presence has any effect on the inducibility of the phages themselves. The phage genomes in *L. lactis* SK11 and *L. lactis* IL1403 are not interrupted by IS elements, except for phage bIL311 in *L. lactis* IL1403, which contains two IS*983* elements (16).

An in-depth analysis of the phages from all three strains will be presented elsewhere (M. Ventura, A. Zomer, C. Canchaya, O. P. Kuipers, J. Kok, and D. van Sinderen, unpublished data).

Integration hot spot and chromosomal inversion. A 59-kb DNA region (nucleotides 647,000 to 706,000) contains genes that have been reported to be plasmid encoded, e.g., two remnants of plasmid replication protein genes, genes of the HsdRMS restriction/modification system, and the *opp-pepO* operon (96). This region also contains three repeat sequences (Fig. 2), one of which is highly similar to a binding site for a recombinase of the resolvase family (PFAM PF00239) and the resolvase gene itself (*tnpR*). The homologous resolvase (CAA36950.1) in *Staphylococcus aureus* is involved in integration and excision of DNA (78). The repeat sequences are specifically found on plasmids and other bacterial genomes, including the "lateral gene transfer hot spot" in the genome of *Streptococcus thermophilus* (9). A number of such integration (and subsequent partial deletion) events could explain the large number of plasmid genes in this area of the genome. The parental strain of *L. lactis* MG1363, *L. lactis* NCDO712 (23), contains only one chromosomal *opp* gene cluster, coding for the oligopeptide transport system. This cluster is located in the same position as the *opp-2* genes in *L. lactis* MG1363. In addition, strain NCDO712 contains an *opp-pepO* gene cluster on a plasmid of at least 40 kb (60). The chromosomally located *opp-2* system in *L. lactis* MG1363 is inactive, since some of the *opp-2* genes contain frameshift mutations (81). From the genome comparison with *L. lactis* IL1403, it can be concluded that *opp-2* is the original copy of this locus. In *L. lactis* MG1363, the *oppD1* promoter is regulated by CodY (24). A clustalW analysis of the upstream region of *oppD1*, *oppD2*, and *oppD* of *L. lactis* IL1403 showed marked differences (data not shown). Interestingly, the CodY binding site is missing in the region upstream of *oppD2*, as well as in *oppD* in *L. lactis* IL1403, indicating that *L. lactis* MG1363 and *L. lactis* IL1403 should differ with respect to the regulation of their active *opp* systems. It is noteworthy that an *oppA1* mutant is not able to grow on the peptide Leu-enkephalin (96), showing that the apparently intact *oppA2* gene does not complement this mutation. This could be due to a lack of transcription of the respective gene or the fact that OppA1 and OppA2, although the same size, show only 88% sequence identity. The fortuitous insertion, during plasmid curing of *L. lactis* NCDO712 (41), of the plasmid-derived *opp* genes in the integration hot spot gives the resulting plasmid-free strain, *L. lactis* MG1363, a growth advantage in mixed cultures in milk: although the strain cannot produce oligopeptides from milk casein due to the lack of proteinase (caseinase) activity, it can efficiently utilize the oligopeptides produced by proteinase-positive strains in the starter culture.

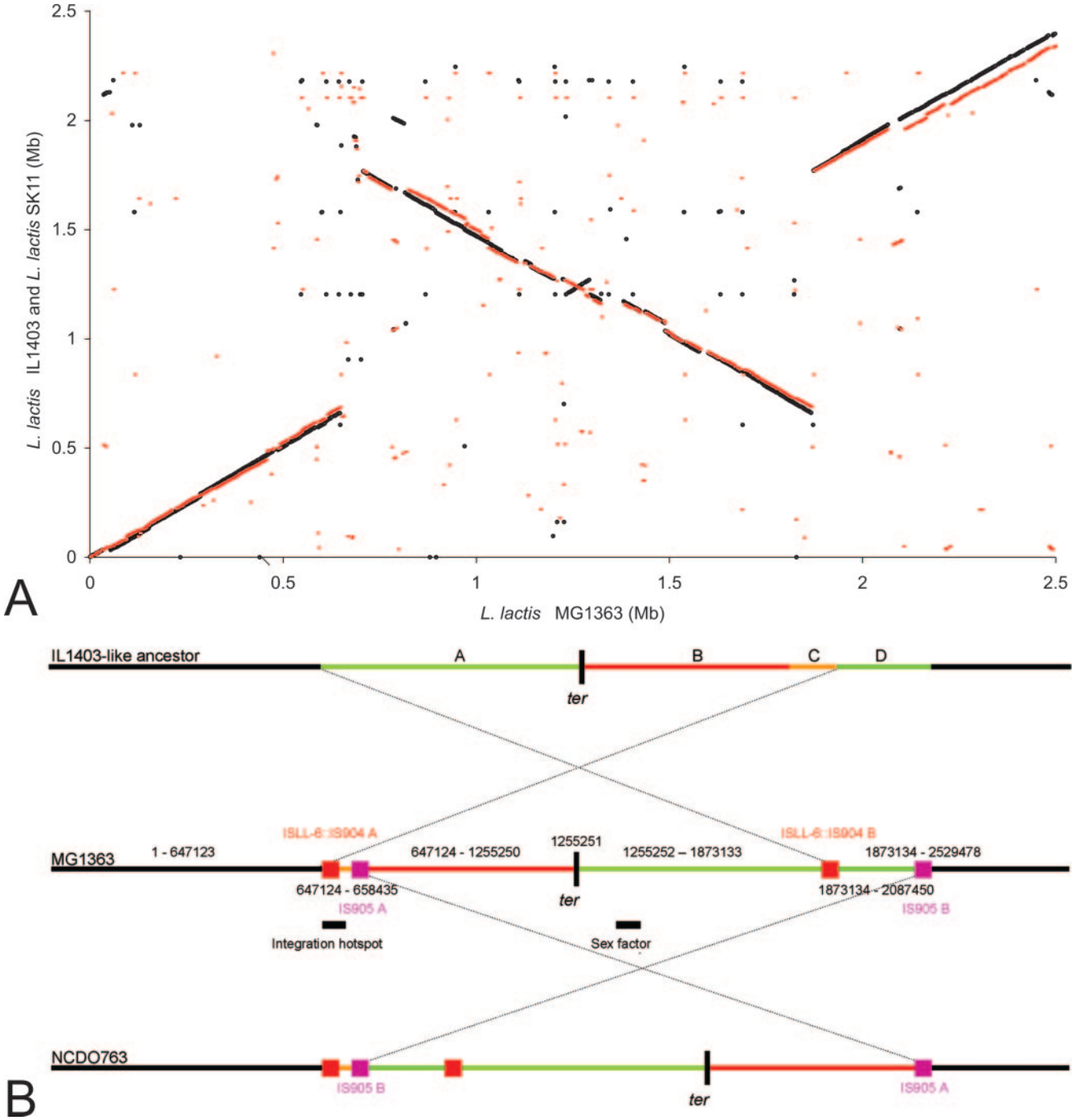

FIG. 3. Chromosomal inversions in *L. lactis*. (A) Dot plot of the nucleotide sequences of the chromosomes of *L. lactis* IL1403 (red) and *L. lactis* SK11 (black) (both on the *y* axis) against that of *L. lactis* MG1363 (*x* axis). (B) Chromosomal inversions in *L. lactis* MG1363 and *L. lactis* NCDO763. The colored squares indicate homologous sequence pairs involved in the inversions, while the numbers indicate the coordinates in the *L. lactis* MG1363 genome sequence of the fragments involved. Horizontal black bars show the positions of the integration hot spot and the sex factor. *ter*, putative terminus of replication. See the text for details.

The chromosomes of *L. lactis* MG1363, *L. lactis* SK11, and *L. lactis* IL1403 show extensive gene synteny when a large chromosomal inversion, described previously (61), is taken into account (Fig. 3A). The *L. lactis* MG1363 integration hot spot is located exactly at the left-hand end of this inversion region. The inversion occurs between the nucleotide positions 647123 and 1873134 and, with a total of 1,226,011 bp, represents 48% of the genome. The flanking regions in the three strains are identical, suggesting that the inversion in *L. lactis* MG1363 took place relatively recently, long after the subspecies division. *L. lactis* SK11 has the same genome organization as *L. lactis* IL1403, with the exception of a small 73-kb inversion

near *terC* (see below). Analysis of the genetic loci at the left and right borders of the large inversion in *L. lactis* MG1363 (Fig. 3B) suggests that the following sequence of events has led to the inversion. (i) A copy of IS*904* transposed into a copy of IS*LL6* in one of the border locations. (ii) A subsequent duplication and transposition of this region occurred, most likely involving a plasmid-encoded IS*LL6* transposase, as there is no functional copy of IS*LL6* left on the chromosome of *L. lactis* MG1363. (iii) Finally, a flip between the two divergently oriented duplicated sequences took place through homologous recombination. The inversion disturbs the symmetry between *oriC* and *ter* only marginally, by shifting *terC* a further 9.4 kb away from the symmetry center (Fig. 1). Since several PCR and Southern hybridization approaches failed to show the reversal of the 1.2-Mb segment, it is assumed that the inversion is stable and that *L. lactis* MG1363 cells in a culture all carry a genome with the structure published in this paper (data not shown). A similar inversion event was described previously (22) for *L. lactis* NCDO763, another derivative of *L. lactis* NCDO712 (60). This event concerned another inactive IS element, one that is also located in the integration hot-spot region of the *L. lactis* MG1363 chromosome. In this case, two inactivated copies of IS*905* were involved, and the subsequent inversion comprised 56% of the genome. The inversion in NCDO763 results in the reinstatement of the original IL1403-like orientation of the central regions (regions A and B in Fig. 3B) on either side of *ter*. The left-hand border region (region C) between IS*LL6*::IS*904A* and IS*905A*, as seen in *L. lactis* MG1363, remains in place, and the DNA region (region D), originally to the right of the *L. lactis* MG1363 inversion region between the IS*LL6*::IS*904A* composite transposon and the IS*905B* transposon, has now been inverted and transposed next to region C. In *L. lactis* NCDO763, the DNA rearrangement has resulted in moving the *ter* site 200 kb to the right, affecting the symmetry between the origin and the terminus loci. The right side of the inversion in *L. lactis* MG1363 around position 1867516 contains the right junction, as well as the *xis*, *int*, and *ardA* genes of transposon Tn*5276*, indicating that this transposon has been deleted from the chromosome (Fig. 3B). Since the orientations of the CDSs of *L. lactis* MG1363 are similar to those of its parent, *L. lactis* NCDO712 (60), the inversion must have occurred before the latter strain was deposited in the National Collection of Dairy Organisms, Reading, United Kingdom in 1954 (23) and might have resulted from the integration of foreign or plasmid DNA. This implies that the integration hot spot was not caused by the plasmid curing of *L. lactis* NCDO712 during the process of *L. lactis* MG1363 construction. The hot-spot region, together with the conjugative sex factor, may have provided the ancestor of *L. lactis* MG1363 with flexibility in the uptake and maintenance of DNA.

A 73-kb chromosomal inversion in *L. lactis* SK11 relative to *L. lactis* IL1403 and *L. lactis* MG1363 is present in a region near *terC*, between the *L. lactis* SK11 nucleotide positions 1202292 and 1275464 (Fig. 3A). This inversion has been facilitated through homologous recombination between two divergently oriented IS*981* elements. This region in the three lactococcal strains contains genes usually found on plasmids: an integrase-recombinase gene (PF00589) and various IS elements. In *L. lactis* IL1403, it carries the citrate utilization genes. The region in *L. lactis* MG1363 contains a duplication of *cspD* (see below); two copies of *ceo*, a gene that is also present on the sucrose transposon Tn*5306* (30); an additional recombinase-encoding gene; and a putative arsenic resistance gene cluster. The region in *L. lactis* SK11 is very similar to that of *L. lactis* MG1363, but it lacks some of the genes present in *L. lactis* MG1363. The presence of these and other possibly plasmid-derived sequences is suggestive of a plasmid insertion in the ancestor of all three *Lactococcus* strains, followed by differential reductive evolution events.

Correlation of genotypes with the classical lactococcal subspecies phenotypes. *L. lactis* strains used in the dairy industry are divided into the two subspecies *lactis* and *cremoris*. Historically, these have been distinguished based on (industrially relevant) phenotypic properties, such as the ability of *L. lactis* subsp. *lactis* strains to metabolize arginine and maltose and to grow at 40°C and in the presence of 4% NaCl, in contrast to the *L. lactis* subsp. *cremoris* strains (95). Thus, *L. lactis* MG1363 was classified initially as *L. lactis* subsp. *lactis*. More recently, on the basis of thorough molecular-genetic analyses, including DNA-DNA hybridization, 16S rRNA and gene sequencing, and PCR-based typing methods, *L. lactis* MG1363 was reclassified as *L. lactis* subsp. *cremoris*. However, it represents an atypical example of the subspecies, as, e.g., it contains an active ADI pathway for arginine degradation (44, 98). In this respect, *L. lactis* subsp. *cremoris* SK11 is a "true" *cremoris* strain. The arginine deiminase-negative phenotype of *L. lactis* SK11, however, is attributable to a single-base-pair deletion at position 675 of the *arcA* gene, creating a frameshift and subsequent pseudogene in *L. lactis* SK11. Orthologs of all other ADI pathway genes are present in *L. lactis* SK11. The use of maltose as a carbon source can be correlated with the genomic sequences of the *L. lactis* strains MG1363, IL1403, and SK11 (see below). Other phenotypic traits are more difficult to identify from the genome, as they are multifactorial.

Carbon utilization. Forty-seven of the genes present in *L. lactis* MG1363, but not in *L. lactis* IL1403, belong to the COG functional category "carbohydrate metabolism and transport" (94), by far the largest category of novel genes in this strain relative to *L. lactis* IL1403. Seventeen genes of this category are unique in *L. lactis* MG1363 compared to *L. lactis* SK11 (Table 4). Both *L. lactis* MG1363 and *L. lactis* IL1403 were tested on Biolog Phenotype MicroArray plates for carbon catabolism (7). The results obtained suggest that *L. lactis* MG1363 has a much greater capability of growing on various sugars, especially those found in plant material (Fig. 4 and Table 4). For example, polyols, like mannitol and sorbitol, are utilized by *L. lactis* MG1363 but not by *L. lactis* IL1403, which has a frameshift in *mtlA*, encoding the mannitol-specific enzyme II component of the phosphotransferase system (PTS). Conversely, dye reduction on γ-cyclodextrine and glycerol was higher for *L. lactis* IL1403 than for *L. lactis* MG1363. Indeed, *L. lactis* MG1363 contains a frameshift in *glpF1*, encoding a putative glycerol uptake facilitator. Although both strains contain a second glycerol uptake facilitator gene, *glpF2*, these results indicate that GlpF2 expression or activity might not be sufficient for growth on glycerol of *L. lactis* MG1363. The difference seen in γ-cyclodextrine usage cannot be fully explained. The *L. lactis* MG1363 genome does contain putative genes for the breakdown of polysaccharides with 1,4-glucosidic bonds, such as maltose and cyclodextrin. The corresponding

TABLE 4. Genes present in *L. lactis* MG1363, but not in *L. lactis* IL1403 or *L. lactis* SK11, belonging to the COG functional category "carbohydrate metabolism and transport"

Locus tag	Gene	Presence in *L. lactis* strain[a]:		(Putative) function
		SK11	IL1403	
llmg_0022	*mtlA*	+	−	PTS system, mannitol-specific IIBC component
llmg_0070		−	−	Putative permease
llmg_0158	*glgB*	+	−	1,4-Alpha-glucan branching enzyme
llmg_0246		+	−	Conserved hypothetical protein
llmg_0247		+	−	Putative UDP-glucose 4-epimerase
llmg_0330		+	−	Putative permease protein
llmg_0423	*drrA*	+	−	Daunorubicin resistance ABC transporter
llmg_0487		−	−	Putative trehalose/maltose hydrolase
llmg_0488		−	−	Multiple sugar-binding protein precursor
llmg_0489		−	−	Sugar transport system permease protein
llmg_0490		−	−	Sugar transport system permease protein
llmg_0666	*oxlT*	+	−	Oxalate/formate antiporter
llmg_0737	*malG*	−	+	Maltose ABC transporter permease protein
llmg_0738	*malF*	−	+	Maltose transport system permease protein
llmg_0957	*rpe-2*	+	−	Ribulose-phosphate 3-epimerase
llmg_0958	*rpiB*	+	−	Ribose 5-phosphate isomerase B
llmg_0959		+	−	Beta-glucosidase
llmg_0960		+	−	Beta-glucosidase
llmg_0961		+	−	Sugar kinase and transcriptional regulator
llmg_0963		+	−	PTS system, IIC component
llmg_1009	*lplB*	+	−	Sugar ABC transporter substrate binding protein
llmg_1010	*lplC*	+	−	Sugar ABC transporter permease
llmg_1011	*lplA*	+	−	Sugar ABC transporter substrate-binding protein
llmg_1162		+	−	Sialic acid-specific 9-*O*-acetylesterase
llmg_1163		+	−	Sugar ABC transporter substrate binding
llmg_1164		+	−	Putative membrane protein
llmg_1165		+	−	Putative membrane protein
llmg_1166		+	−	Putative endoglucanase
llmg_1168		+	−	Putative polysaccharide deacetylase
llmg_1169	*aguA*	+	−	Alpha-glucuronidase
llmg_1241		−	−	Conserved hypothetical protein
llmg_1242		−	−	Conserved hypothetical protein
llmg_1244		−	−	Conserved hypothetical protein
llmg_1320		+	−	Putative xylan beta-1,4-xylosidase
llmg_1321		+	−	Conserved hypothetical protein
llmg_1322		+	−	Conserved hypothetical protein
llmg_1358	*orf48*	−	−	Conserved hypothetical protein
llmg_1454		−	−	Putative sugar kinase
llmg_1455	*bglA2*	−	−	6-Phospho-beta-glucosidase
llmg_1456	*bglX*	−	−	Beta-glucosidase
llmg_1608		−	−	Putative glycosyl hydrolases.
llmg_1622	*tagG*	−	+	Teichoic acid ABC transporter permease protein
llmg_1623	*tagH*	−	+	Teichoic acid export ATP-binding protein
llmg_1838		+	−	Putative sugar ABC transporter
llmg_1839		+	−	Putative sugar ABC transporter permease
llmg_2003	*galE*	+	−	UDP-glucose 4-epimerase
llmg_2237	*galP*	+	−	Galactose permease
llmg_2279	*bcrA*	+	−	Putative bacitracin transport ATP binding
llmg_2346	*gmhA*	+	−	Phosphoheptose isomerase
llmg_2439		+	−	Conserved hypothetical protein
llmg_2467	*gntP*	+	−	Gluconate transport protein

[a] +, present; −, absent.

region in the chromosome of *L. lactis* IL1403 contains a number of hypothetical genes not present in *L. lactis* MG1363 that may explain the observed differences between the two strains. *L. lactis* SK11 does not ferment maltose. As Table 4 shows, this strain lacks *malG* and *malF*, genes encoding the maltose permease.

L. lactis strains MG1363, SK11, and IL1403 all carry genes encoding the potential lactate dehydrogenases LdhA, LdhB, and LdhX and a malate/lactate dehydrogenase (HicD in *L. lactis* IL1403). *L. lactis* MG1363 and *L. lactis* SK11 do not contain the *cit* operon for citrate utilization, which is present in *L. lactis* IL1403. *L. lactis* MG1363, like *L. lactis* IL1403, contains an incomplete tricarboxylic acid cycle and carries several genes necessary for aerobic respiration. Unlike *L. lactis* MG1363 and *L. lactis* IL1403, the genes for citrate synthetase (*gltA*), aconitase (*citB*), and malolactic enzyme (*mleS*) are pseudogenes in *L. lactis* SK11. All three lactococcal strains have the *men* and *cytABCD* operons (menaquinone synthesis and cytochrome *d* biogenesis) and the *hemH*, *hemK*, and *hemN* genes for the late steps of heme synthesis. Duwat et al., Rezaïki

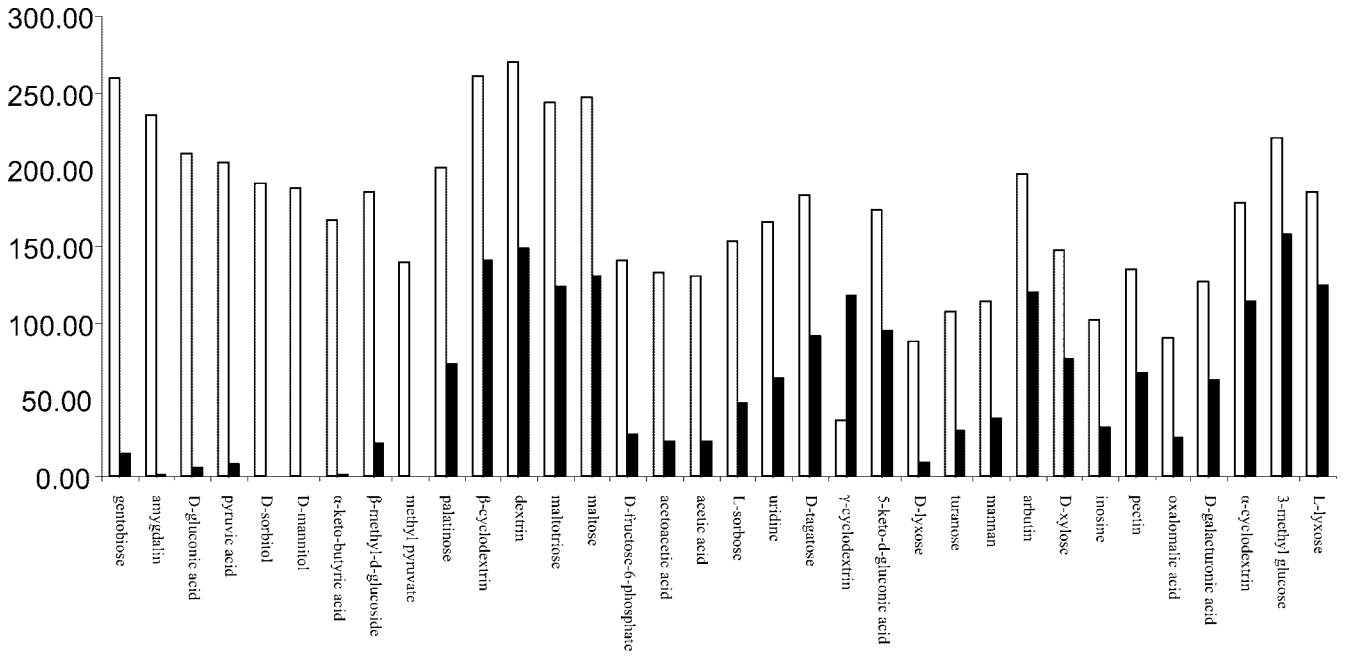

FIG. 4. Growth of *L. lactis* on sugars. Utilization by *L. lactis* MG1363 (open bars) and *L. lactis* IL1403 (black bars) of various sugar sources as measured by a significant (<60) difference in dye intensity development (indicated in arbitrary units on the *y* axis) on Biolog plates (7).

et al., and Vido et al. (34, 77, 102) have demonstrated that *L. lactis* MG1363 is capable of biphasic fermentative and respiratory modes of growth under respiratory conditions, leading to a greater growth yield and better long-term survival. Their physiological and biochemical analyses confirm the *L. lactis* MG1363 genomic information.

There are many indications that the ancestor of *L. lactis* MG1363 occupied a plant-associated niche. *L. lactis* MG1363 still retains the ability to metabolize plant-derived sugars as carbon sources and has cell surface proteins (*csc* clusters of CscA, -B, -C, and -D) that are present in other gram-positive bacteria found in a plant environment (see "The lactococcal cell envelope and secretome" below). Adaptation to milk has included changes in metabolic activity, including the inactivation of several of the amino acid biosynthetic pathways (100). The ancestor of *L. lactis* MG1363 was obviously flexible in its ability to acquire DNA from other bacteria, in particular, plasmids and other mobile elements (see above). This property has contributed to the organism's capacity to adapt from a plant-associated niche to survival in the milk environment, e.g., through the acquisition of plasmids specifying the proteinase (caseinase) PrtP and enzymes and transport proteins for the utilization of the milk sugar lactose (43).

Amino acid-, vitamin-, and nucleic acid biosynthesis. The six amino acids glutamate, leucine, isoleucine, valine, histidine, and methionine are essential for the growth of *L. lactis* (74, 76), while *L. lactis* MG1363 grows to an appreciable extent only when, in addition, one of the five amino acids asparagine, glutamine, alanine, arginine, and threonine is present (53). The many conflicting results on the minimal growth requirements of *L. lactis* may be attributed to the use of different formulations of chemically defined medium (CDM), as concentrations of CDM constituents, such as amino acids, can affect the growth rates of different *L. lactis* strains (44). For example, excess isoleucine results in blocking of several CodY-dependent amino acid biosynthetic pathways and subsequent growth inhibition (45). The *L. lactis* MG1363 genome contains homologues of genes with the potential for biosynthesis of five of the six essential amino acids. In both *L. lactis* MG1363 and *L. lactis* IL1403, *leuA* and *leuB* (leucine biosynthesis) are pseudogenes, while in *L. lactis* SK11, the genes are not mutated. Further differences between *L. lactis* MG1363, *L. lactis* SK11, and *L. lactis* IL1403 exist in the predicted biosynthetic pathways for the nonessential amino acids. *L. lactis* MG1363 harbors a frameshift mutation in the gene for glycerate kinase, the enzyme catalyzing the first step in serine biosynthesis from glycerate. A potential alternative route for serine biosynthesis by *L. lactis* MG1363 is from pyruvate. Threonine biosynthesis can potentially occur from aspartate or by condensation of glycine by threonine aldolase. Jensen et al. (52) observed in vivo activity of threonine aldolase in *L. lactis* MG1363, while 95% of the threonine was taken up from the medium. Growth of *L. lactis* IL1403 was reduced by 25% in the absence of threonine (19). A threonine aldolase gene is not present in the genome of *L. lactis* MG1363, *L. lactis* SK11, or *L. lactis* IL1403, implying that all threonine biosynthesis occurs through aspartate or that a cryptic gene specifies the activity measured by Jensen and coworkers (52). The gene encoding histidinol-phosphate aminotransferase, responsible for the conversion of phenylpyruvate or 4-hydroxy-phenylpyruvate to tyrosine and phenylalanine, respectively, is present in *L. lactis* MG1363 and *L. lactis* SK11 but absent in *L. lactis* IL1403. In the last strain, formation of phenylalanine and tyrosine is predicted to be catalyzed by alternative routes involving histidinol-phosphate aminotransferase and prephenate dehydratase, respectively. Lysine is synthesized from aspartate; clear homologues of the

TABLE 5. Regulators present in *L. lactis* subsp. *cremoris* MG1363 but absent in *L. lactis* subsp. *cremoris* SK11 or *L. lactis* subsp. *lactis* IL1403

Locus tag	Gene name	Presence in *L. lactis* strain[a]:		Annotation	Family
		SK11	IL1403		
llmg_0069		−	−	HTH-type transcriptional regulator	Rgg
llmg_0161		+	−	Putative HTH-type transcriptional regulator	TetR
llmg_0163	*epsR*	+	−	Transcriptional regulator	EpsR
llmg_0301		+	−	Transcriptional regulator	MerR
llmg_0390	*rlrG*	−	+	Transcriptional regulator	LysR
llmg_0424		+	−	Transcriptional regulator	
llmg_0432		−	−	Similar to transcription regulator	DeoR
llmg_0484		−	−	Transcriptional regulator	LacI
llmg_0522		−	+	Putative transcriptional regulator	RpiR
llmg_0572		+	−	Putative transcriptional regulator	MerR
llmg_0603		+	−	HTH-type transcriptional regulator	
llmg_0694	*flpB*	−	−	Transcriptional regulator, FNR like protein B	FNR
llmg_0709		−	−	Transcriptional regulator	PadR
llmg_0865		+	−	Transcriptional antiterminator	BglR
llmg_0925		−	+	Putative transcriptional regulator	EpsR
llmg_0956		+	−	Transcriptional regulator	LacI
llmg_0961		+	−	Sugar kinase and transcriptional regulator	ROK
llmg_0962		+	−	Transcriptional regulator	AraC
llmg_1027	*rmaC*	−	+	Transcriptional regulator	MarR
llmg_1141		−	−	Putative transcriptional regulator	EpsR
llmg_1143		−	−	Hypothetical protein	
llmg_1147		−	−	Putative transcription regulator	
llmg_1209	*rmaX*	−	+	Transcriptional regulator	MarR
llmg_1246	*arsR*	−	−	Regulator of arsenic resistance	ArsR
llmg_1324		+	−	Transcriptional regulator	AraC
llmg_1462		−	−	Putative HTH-type transcriptional regulator	RpiR
llmg_1527		+	−	Putative transcription regulator	
llmg_1627	*rmaH*	−	+	Transcriptional regulator	MarR
llmg_1868		+	−	HTH-type transcriptional regulator	
llmg_1903		+	−	Putative transcriptional regulator	
llmg_2067	*rlrB*	−	+	Transcriptional regulator	LysR
llmg_2334		+	−	Transcriptional regulator	LysR

[a] +, present; −, absent.

required aminotransferase (DapC) and the epimerase (DapF) are present in *L. lactis* MG1363 and *L. lactis* SK11 but absent in *L. lactis* IL1403. The absence in *L. lactis* IL1403 is unusual, since the intermediates of lysine biosynthesis are required for cell wall biosynthesis. Alanine racemase (Alr) may provide the necessary building blocks in *L. lactis* IL1403 and, for that matter, in *L. lactis* MG1363 and *L. lactis* SK11.

In silico analysis of the *L. lactis* MG1363 genome illustrates that nicotinate, pantothenate, biotin, pyridoxine, ubiquinone, and cobalamin cannot be synthesized by this strain, while it remains unclear if thiamine can be synthesized: putative enzymes for synthesis of thiamine phosphate from pyruvate are encoded by the genome of *L. lactis* MG1363, but a gene encoding thiamine kinase, necessary for the conversion of thiamine phosphate to thiamine, could not be identified. Both riboflavin and folic acid are not essential, as riboflavin/FMN and FAD are synthesized from GTP and folic acid from purine metabolism and phenylalanine biosynthesis (15, 93). Folic acid and vitamin B_2 function as cofactors in the synthesis of purines and pyrimidines, and their presence in minimal medium lacking nucleic acids (19, 53) may be essential or stimulatory.

The *L. lactis* MG1363 genome carries genes encoding all enzymes required for the biosynthesis and metabolism of all purines and pyrimidines.

Gene regulation in *L. lactis* MG1363. The genome of *L. lactis* MG1363 encodes 136 proteins with possible functions in gene regulation. Of these, 10 are unique to *L. lactis* MG1363 compared to *L. lactis* IL1403 or *L. lactis* SK11 (Table 5) Seven of the eight two-component regulatory systems (2CSs) present in *L. lactis* MG1363 have orthologs in *L. lactis* SK11 and *L. lactis* IL1403. The eighth 2CS (llmg_pseudo_52 and llmg_pseudo_53), which is absent from the *L. lactis* IL1403 genome, harbors frameshift mutations in both the histidine kinase and response regulator genes in *L. lactis* MG1363 and is unlikely to be functional in this strain. This 2CS is intact in *L. lactis* SK11.

The lactococcal cell envelope and secretome. *L. lactis* MG1363 is predicted to secrete around 184 (7.5%) proteins. Of these, 39 contain a consensus lipoprotein signal peptide and are expected to be attached to the cytoplasmic membrane, while 38 are (putative) cell wall-attached proteins, as they contain a cell wall-anchoring motif(s) leading to covalent (9 proteins) or noncovalent (29 proteins) attachment. Of the LPXTG motif-containing proteins of *L. lactis* IL1403 (12 proteins) and *L. lactis* MG1363 (9 proteins), 4 are common to both strains. *L. lactis* MG1363, *L. lactis* SK11, and *L. lactis* IL1403 all contain two putative sortase homologues, SrtA and SrtC, which are assumed to be distinguishable on the basis of their substrate specificities, namely, cleavage between the threonine

TABLE 6. Glycosyltransferase gene clusters present in *L. lactis* MG1363 but not in IL1403

Locus tag	% GC	Best BLASTP hit	Accession no.	Amino acid identity (%)
Cluster 1				
llmg_0220	28.6	*S. thermophilus* CNRZ1066	gb AAV63009.1	49
llmg_0221	29.7	*S. pneumoniae*	emb CAI33818.1	31
llmg_0223	29	*S. pneumoniae*	emb CAI34546.1	32
Cluster 2				
llmg_1617	36.2	*E. faecalis* V583	gb AAO82207.1	33
llmg_1620	38.9	*Enterococcus faecalis* V584	gb AAO82208.1	43
llmg_1621	35.6	*E. faecalis* V585	gb AAO82208.1	55
Cluster 3				
llmg_1707	33	*S. agalactiae* 2603V/R	gb AN00323.1	57
llmg_1708	29.2	*Lactobacillus johnsonii* NCC533	gb AAS08043.1	21
llmg_1710	33.2	*L. johnsonii* NCC533	gb AAS08382.1	46
Cluster4				
llmg_2344	34.5	*S. agalactiae*	gb AAK43612.1	40
llmg_2349	31.9	*S. thermophilus*	gb AAV60736.1	28
llmg_2350	31	*Bacillus fragilis* NCTC 9343	emb CAH09494.1	37
llmg_2351	31.9	*S. thermophilus*	gb AAN63513.1	31

and glycine residues in the sequence LP(KENQA)TG(EDS) (SrtA) or (LI)P(KSELAN)TG(GVTSA) (SrtC) (20). Interestingly, the LPXTG motif-containing proteins unique to *L. lactis* MG1363 are all SrtA targets.

Recently, a gene cluster typically comprising the genes *cscABCD*, which is present only in a subgroup of gram-positive bacteria, has been described to be possibly related to association with plants (84). Four *csc* clusters are present in *L. lactis* MG1363, three of which are also present in *L. lactis* IL1403. The same four clusters are present in the *L. lactis* SK11 chromosome, while this strain carries an additional *csc* operon on one of its plasmids. The cluster extending from llmg_1503 to llmg_1507, which is not present in *L. lactis* IL1403, is involved in UV sensitivity in *L. lactis* MG1363 (33).

There has been a growing interest in the food industry in extracellular polysaccharide-producing LAB because of their influence on the textural characteristics of fermented food products. Although 28 putative glycosyltransferases are encoded by the *L. lactis* MG1363 genome, we could not identify a gene cluster with the typical genetic organization of a lactococcal extracellular polysaccharide operon (21). The glycosyltransferases are more likely to be involved in the decoration of the lactococcal cell wall with cell wall polysaccharides (WPS). In comparison with *L. lactis* IL1403, *L. lactis* MG1363 harbors at least 13 unique glycosyltransferase genes, suggesting that distinct differences exist in the compositions of the WPS of these two strains. These unique genes are organized in three clusters of three glycosyltransferase genes and one cluster of four genes (Table 6). One cluster displays a distinctly lower GC content (29%) than that of the average *L. lactis* DNA, indicating an HGT event.

Recently, it has been demonstrated that the receptor binding proteins of lactococcal phages bind to WPS structures in the bacterial cell wall (31). On the basis of the C-terminal parts of their receptor binding proteins, the bacteriophages investigated could be grouped into *L. lactis* subsp. *lactis*- and *L. lactis* subsp. *cremoris*-infecting phages (32). Today, our understanding of the function and physiological relevance of WPS for the bacterial cell is very limited, but one of the obvious effects of distinct sets of glycosyltransferases may be the different susceptibilities of the bacteria to certain bacteriophages.

Conclusions. Despite the fact that *L. lactis* IL1403 carries 41 kb more bacteriophage-related sequences, the *L. lactis* MG1363 genome is significantly bigger (160 kb) than that of *L. lactis* IL1403. This is largely due to the presence of mobile genetic elements in *L. lactis* MG1363, namely, the unique sex factor and a larger number of insertion elements, and the integration hot-spot region. The latter chromosomal section enables *L. lactis* MG1363 to stably integrate laterally acquired DNA, as is documented by the large number of plasmid-related genes therein. Our findings show that the integration hot spot has played a key role in the evolution of the genome of *L. lactis* MG1363 and its related strains, influencing the genome content, as well as its overall structure, as demonstrated by the insertion sequence-mediated chromosomal inversion. It allowed *L. lactis* MG1363, e.g., to stably maintain a functional copy of the *opp* operon, an operon that is essential for growth in milk, thus contributing to the overall "fitness" of the strain. Forty-seven of the genes present in *L. lactis* MG1363 but absent in *L. lactis* IL1403 belong to the COG functional category "carbohydrate metabolism and transport." Consequently, *L. lactis* MG1363 displays a greater capability to grow on various sugars, especially those found in plant material. The ability to metabolize plant-derived sugars and the fact that the *L. lactis* MG1363 genome carries four *csc* gene clusters associated with the utilization of complex plant polysaccharides point to a plant-associated biological niche for the ancestor of *L. lactis* MG1363.

The *L. lactis* MG1363 genome sequence presented here is vital to both applied and fundamental research into LAB and other low-percent GC gram-positive bacteria. This is especially important in view of current and future uses of *L. lactis* MG1363 in medical research, which today includes the delivery in the gastrointestinal tract of bioactive molecules and, impor-

tantly, the therapeutic use of a live *L. lactis* MG1363 derivative strain in a human trial study. As we now have the genome sequences of both *Lactococcus* subspecies, *L. lactis* subsp. *cremoris* strain MG1363, reported here, and *L. lactis* subsp. *lactis* IL1403 (11), and, very recently, *L. lactis* subsp. *cremoris* SK11 (66)—and with several genomes of other, plant-derived, strains to come (98)—these are promising times for in-depth comparative-genomics analyses of this economically important bacterial species.

ACKNOWLEDGMENTS

Udo Wegmann, Michael J. Gasson, and Claire Shearman acknowledge funding by a CSG grant from the BBSRC research council. Mary O'Connell-Motherway and Douwe van Sinderen acknowledge the Higher Education Authority and Science Foundation Ireland for financial support (grants PRTLI Cycle 3 and 04/BRB0647, respectively). Carlos Canchaya is supported by an EMBARK postdoctoral fellowship. Aldert Zomer and Girbe Buist acknowledge funding by IOP Genomics grant IGE01018.

Jan Kok, Oscar P. Kuipers, Girbe Buist, and Aldert Zomer acknowledge Sacha van Hijum and Anne de Jong for their support and helpful discussions. Mary O'Connell-Motherway and Douwe van Sinderen acknowledge Sinead Leahy and Gerald Fitzgerald for their support and assistance.

The groups of Michael J. Gasson, Jan Kok/Oscar P. Kuipers, and Douwe van Sinderen contributed equally to the initiation, financing, and execution of these studies.

REFERENCES

1. **Altermann, E., W. M. Russell, M. A. Azcarate-Peril, R. Barrangou, B. L. Buck, O. McAuliffe, N. Souther, A. Dobson, T. Duong, M. Callanan, S. Lick, A. Hamrick, R. Cano, and T. R. Klaenhammer.** 2005. Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM. Proc. Natl. Acad. Sci. USA **102:**3906–3912.
2. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
3. **Audouy, S. A. L., M. L. van Roosmalen, J. Neef, R. Kanninga, E. Post, M. van Deemter, H. Metselaar, S. van Selm, G. T. Robillard, K. J. Leenhouts, and P. W. M. Hermans.** 2006. *Lactococcus lactis* GEM particles displaying pneumococcal antigens induce local and systemic immune responses following intranasal immunization. Vaccine **24:**5434–5441.
4. **Badger, J. H., and G. J. Olsen.** 1999. CRITICA: coding region identification tool invoking comparative analysis. Mol. Biol. Evol. **16:**512–524.
5. **Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. L. Sonnhammer.** 2002. The Pfam protein families database. Nucleic Acids Res. **30:**276–280.
6. **Bermudez-Humaran, L. G., N. G. Cortes-Perez, F. Lefevre, V. Guimaraes, S. Rabot, J. M. Alcocer-Gonzalez, J.-J. Gratadoux, C. Rodriguez-Padilla, R. S. Tamez-Guerra, G. Corthier, A. Gruss, and P. Langella.** 2005. A novel mucosal vaccine based on live Lactococci expressing E7 antigen and IL-12 induces systemic and mucosal immune responses and protects mice against human Papillomavirus type 16-induced tumors. J. Immunol. **175:**7297–7302.
7. **Bochner, B. R., P. Gadzinski, and E. Panomitros.** 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res. **11:**1246–1255.
8. **Boeckmann, B., A. Bairoch, R. Apweiler, M.-C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, and M. Schneider.** 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. **31:**365–370.
9. **Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek, N. Kyprides, B. Purnelle, D. Prozzi, K. Ngui, D. Masuy, F. Hancy, S. Burteau, M. Boutry, J. Delcour, A. Goffeau, and P. Hols.** 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. Nat. Biotechnol. **22:**1554–1558.
10. **Bolotin, A., B. Quinquis, A. Sorokin, and D. S. Ehrlich.** 2004. Recent genetic transfer between *Lactococcus lactis* and enterobacteria. J. Bacteriol. **186:**6671–6677.
11. **Bolotin, A., P. Wincker, S. Mauger, O. Jaillon, K. Malarme, J. Weissenbach, S. D. Ehrlich, and A. Sorokin.** 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. Genome Res. **11:**731–753.
12. **Bongers, R. S., M. H. N. Hoefnagel, M. Starrenburg, M. A. J. Siemerink, J. G. A. Arends, J. Hugenholtz, and M. Kleerebezem.** 2003. IS*981*-mediated adaptive evolution recovers lactate production by *ldhB* transcription activation in a lactate dehydrogenase-deficient strain of *Lactococcus lactis*. J. Bacteriol. **185:**4499–4507.
13. **Brown, C. M., M. E. Dalphin, P. A. Stockwell, and P. Warren.** 1993. Tate, the translational termination signal database. Nucleic Acids Res. **21:**3119–3123.
14. **Bruno-Barcena, J. M., M. A. Azcarate-Peril, T. R. Klaenhammer, and H. M. Hassan.** 2005. Marker-free chromosomal integration of the manganese superoxide dismutase gene (*sodA*) from *Streptococcus thermophilus* into *Lactobacillus gasseri*. FEMS Microbiol. Lett. **246:**91–101.
15. **Burgess, C., M. O'Connell-Motherway, W. Sybesma, J. Hugenholtz, and D. van Sinderen.** 2004. Riboflavin production in *Lactococcus lactis*: potential for *in situ* production of vitamin-enriched foods. Appl. Environ. Microbiol. **70:**5769–5777.
16. **Chopin, A., A. Bolotin, A. Sorokin, D. Ehrlich, and M. C. Chopin.** 2001. Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. Nucleic Acids Res. **29:**644–651.
17. **Chopin, M.-C., A. Chopin, and E. Bidnenko.** 2005. Phage abortive infection in lactococci: variations on a theme. Curr. Opin. Microbiol. **8:**473–479.
18. **Christie, J., R. McNab, and H. F. Jenkinson.** 2002. Expression of fibronectin-binding protein FbpA modulates adhesion in *Streptococcus gordonii*. Microbiology **148:**1615–1625.
19. **Cocaign-Bousquet, M., C. Garrigues, L. Novak, N. D. Lindley, and P. Loubiere.** 1995. Rational development of a simple synthetic medium for the sustained growth of *Lactococcus lactis*. J. Appl. Bacteriol. **79:**108–116.
20. **Comfort, D., and R. T. Clubb.** 2004. A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria. Infect. Immun. **72:**2710–2722.
21. **Dabour, N., and G. LaPointe.** 2005. Identification and molecular characterization of the chromosomal exopolysaccharide biosynthesis gene cluster from *Lactococcus lactis* subsp. *cremoris* SMQ-461. Appl. Environ. Microbiol. **71:**7414–7425.
22. **Daveran-Mingot, M.-L., N. Campo, P. Ritzenthaler, and P. Le Bourgeois.** 1998. A natural large chromosomal inversion in *Lactococcus lactis* is mediated by homologous recombination between two insertion sequences. J. Bacteriol. **180:**4834–4842.
23. **Davies, F. L., H. M. Underwood, and M. J. Gasson.** 1981. The value of plasmid profiles for strain identification in lactic streptococci and the relationship between *Streptococcus lactis* 712, ML3 and C2. J. Appl. Bacteriol. **51:**325–337.
24. **den Hengst, C. D., P. Curley, R. Larsen, G. Buist, A. Nauta, D. van Sinderen, O. P. Kuipers, and J. Kok.** 2005. Probing direct interactions between CodY and the *oppD* promoter of *Lactococcus lactis*. J. Bacteriol. **187:**512–521.
25. **de Ruyter, P. G. G. A., O. P. Kuipers, and W. M. de Vos.** 1996. Controlled gene expression systems for *Lactococcus lactis* with the food grade inducer nisin. Appl. Environ. Microbiol. **62:**3662–3667.
26. **de Vos, W. M., and J. Hugenholtz.** 2003. Engineering metabolic highways in lactococci and other lactic acid bacteria. Trends Biotechnol. **22:**72–79.
27. **Dodd, H. M., N. Horn, and M. J. Gasson.** 1994. Characterisation of IS*905*, a new multicopy insertion sequence identified in lactococci. J. Bacteriol. **176:**3393–3396.
28. **Dodd, H. M., N. Horn, Z. Hao, and M. J. Gasson.** 1992. A lactococcal expression system for engineering nisins. Appl. Environ. Microbiol. **58:**3683–3693.
29. **Dodd, I. B., and J. B. Egan.** 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18:**5019–5026.
30. **Donkersloot, J. A., and J. Thompson.** 1995. Cloning, expression, sequence analysis, and site-directed mutagenesis of the Tn*5306*-encoded *N*-(carboxyethyl)ornithine synthase from *Lactococcus lactis* K1. J. Biol. Chem. **270:**12226–12234.
31. **Dupont, K., T. Janzen, F. K. Vogensen, J. Josephsen, and B. Stuer-Lauridsen.** 2004. Identification of *Lactococcus lactis* genes required for bacteriophage adsorption. Appl. Environ. Microbiol. **70:**5825–5832.
32. **Dupont, K., F. K. Vogensen, H. Neve, J. Bresciani, and J. Josephsen.** 2004. Identification of the receptor-binding protein in 936-species lactococcal bacteriophages. Appl. Environ. Microbiol. **70:**5818–5824.
33. **Duwat, P., A. Cochu, S. D. Ehrlich, and A. Gruss.** 1997. Characterization of *Lactococcus lactis* UV-sensitive mutants obtained by IS*S1* transposition. J. Bacteriol. **179:**4473–4479.
34. **Duwat, P., S. Sourice, B. Cesselin, G. Lamberet, K. Vido, P. Gaudu, Y. Le Loir, F. Violet, P. Loubière, and A. Gruss.** 2001. Respiration capacity of the fermenting bacterium *Lactococcus lactis* and its positive effects on growth and survival. J. Bacteriol. **183:**4509–4516.
35. **Dyrlov Bendtsen, J., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340:**783–795.
36. **Eddy, S. R.** 1996. Hidden Markov models. Curr. Opin. Struct. Biol. **6:**361–365.

37. **Edwards, R. A., G. J. Olsen, and S. R. Maloy.** 2002. Comparative genomics of closely related salmonellae. Trends Microbiol. **10:**94–99.
38. **Ewing, B., and P. Green.** 1989. Basecalling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186–194.
39. **Frank, A. C., and J. R. Lobry.** 2000. Oriloc: prediction of replication boundaries in unannotated bacterial chromosomes. Bioinformatics **16:**560–561.
40. **Gasson, M. J.** 1990. *In vivo* genetic systems in lactic acid bacteria. FEMS Microbiol. Rev. **87:**43–60.
41. **Gasson, M. J.** 1983. Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic streptococci after protoplast-induced curing. J. Bacteriol. **154:**1–9.
42. **Gasson, M. J., J.-J. Godon, C. J. Pillidge, T. J. Eaton, K. Jury, and C. A. Shearman.** 1995. Characterization and exploitation of conjugation in *Lactococcus lactis*. Int. Dairy J. **5:**757–762.
43. **Gasson, M. J., and C. A. Shearman.** 2003. Plasmid biology, conjugation and transposition, p. 25–44. *In* B. J. B. Wood and P. J. Warner (ed.), Genetics of lactic acid bacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.
44. **Godon, J.-J., C. Delorme, S. D. Ehrlich, and P. Renault.** 1992. Divergence of genomic sequences between *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*. Appl. Environ. Microbiol. **58:**4045–4047.
45. **Guedon, E., B. Sperandio, N. Pons, S. D. Ehrlich, and P. Renault.** 2005. Overall control of nitrogen metabolism in *Lactococcus lactis* by CodY, and possible models for CodY regulation in Firmicutes. Microbiol. **151:**3895–3909.
46. **Haft, D. H., B. J. Loftus, D. L. Richardson, F. Yang, J. A. Eisen, I. T. Paulsen, and O. White.** 2001. TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res. **29:**41–43.
47. **Hak-Jong, C., M.-J. Seo, J.-C. Lee, C.-I. Cheigh, H. Park, C. Ahn, and Y.-R. Pyan.** 2005. Heterologous expression of human beta-defensin-1 in bacteriocin-producing *Lactococcus lactis* J. Microbiol. Biotechnol. **15:**330–336.
48. **Harris, M. A., et al.** 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. **32:**D258–D261.
49. **Hols, P., M. Kleerebezem, A. N. Schank, T. Ferain, J. Hugenholtz, J. Delcour, and W. M. de Vos.** 1999. Conversion of *Lactococcus lactis* from homolactic to homoalanine fermentation through metabolic engineering. Nat. Biotechnol. **17:**588–592.
50. **Horn, N., A. Fernandez, H. M. Dodd, M. J. Gasson, and J. M. Rodriguez.** 2004. Nisin-controlled production of pediocin PA-1 and colicin V in nisin- and non-nisin-producing *Lactococcus lactis* strains. Appl. Environ. Microbiol. **70:**5030–5032.
51. **Igimi, S., C. H. Ryu, S. H. Park, Y. Sasaki, T. Sasaki, and S. Kumagai.** 1996. Transfer of conjugative plasmid pAMB1 from *Lactococcus lactis* to mouse intestinal bacteria. Lett. Appl. Microbiol. **23:**31–35.
52. **Jensen, N. B. S., B. Christensen, J. Nielsen, and J. Villadsen.** 2002. The simultaneous biosynthesis and uptake of amino acids by *Lactococcus lactis* studied by ^{13}C-labeling experiments. Biotechnol. Bioeng. **78:**11–16.
53. **Jensen, P. R., and K. Hammer.** 1993. Minimal requirements for exponential growth of *Lactococcus lactis*. Appl. Environ. Microbiol. **59:**4363–4366.
54. **Kanehisa, M., S. Goto, S. Kawashima, and A. Nakaya.** 2002. The KEGG databases at GenomeNet. Nucleic Acids Res. **30:**42–46.
55. **Kleerebezem, M., M. M. Beerthuyzen, E. Vaughan, W. M. de Vos, and O. P. Kuipers.** 1997. Controlled gene expression systems for lactic acid bacteria: transferable nisin-inducible expression cassettes for *Lactococcus*, *Leuconostoc*, and *Lactobacillus* sp. Appl. Environ. Microbiol. **63:**4581–4584.
56. **Kleerebezem, M., J. Boekhorst, R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer, R. Tarchini, S. A. Peters, H. M. Sandbrink, M. W. E. J. Fiers, W. Stiekema, R. M. K. Lankhorst, P. A. Bron, S. M. Hoffer, M. N. N. Groot, R. Kerkhoven, M. de Vries, B. Ursing, W. M. de Vos, and R. J. Siezen.** 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Natl. Acad. Sci. USA **100:**1990–1995.
57. **Klein, N., and S. Lortal.** 1999. Attenuated starters: an efficient means to influence cheese ripening—a review. Int. Dairy J. **9:**751–762.
58. **Kloosterman, T. G., J. J. E. Bijlsma, J. Kok, and O. P. Kuipers.** 2006. To have neighbour's fare: extending the molecular toolbox for *Streptococcus pneumoniae*. Microbiology **152:**351–359.
59. **Krough, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580.
60. **Le Bourgeois, P., M.-L. Daveran-Mingot, and P. Ritzenthaler.** 2000. Genome plasticity among related *Lactococcus* strains: identification of genetic events associated with macrorestriction polymorphisms. J. Bacteriol. **182:**2481–2491.
61. **Le Bourgeois, P., M. Lautier, L. Vanderberghe, M. J. Gasson, and P. Ritzenthaler.** 1995. Physical and genetic map of the *Lactococcus lactis* subsp. *cremoris* MG1363 chromosome comparison with that of *Lactococcus lactis* subsp. *lactis* IL1403 reveals a large genome inversion. J. Bacteriol. **177:**2840–2850.
62. **Le Loir, Y., V. Azevedo, S. Oliveira, D. Freitas, A. Miyoshi, L. Bermudez-Humaran, S. Nouaille, L. Ribeiro, S. Leclercq, J. Gabriel, V. Guimaraes, M. Oliveira, C. Charlier, M. Gautier, and P. Langella.** 2005. Protein secretion in *Lactococcus lactis*: an efficient way to increase the overall heterologous protein production. Microb. Cell Factories **4:**2.
63. **Licht, T. R., B. B. Christensen, K. A. Krogfelt, and S. Molin.** 1999. Plasmid transfer in the animal intestine and other dynamic bacterial populations: the role of community structure and environment. Microbiology **145:**2615–2622.
64. **Liu, Y., P. Harrison, V. Kunin, and M. Gerstein.** 2004. Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. Genome Biol. **5:**R64.
65. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.
66. **Makarova, K., A. Slesarev, Y. Wolf, A. Sorokin, B. Mirkin, E. Koonin, A. Pavlov, N. Pavlova, V. Karamychev, N. Polouchine, V. Shakhova, I. Grigoriev, Y. Lou, D. Rohksar, S. Lucas, K. Huang, D. M. Goodstein, T. Hawkins, V. Plengvidhya, D. Welker, J. Hughes, Y. Goh, A. Benson, K. Baldwin, J. H. Lee, I. Diaz-Muniz, B. Dosti, V. Smeianov, W. Wechter, R. Barabote, G. Lorca, E. Altermann, R. Barrangou, B. Ganesan, Y. Xie, H. Rawsthorne, D. Tamir, C. Parker, F. Breidt, J. Broadbent, R. Hutkins, D. O'Sullivan, J. Steele, G. Unlu, M. Saier, T. Klaenhammer, P. Richardson, S. Kozyavkin, B. Weimer, and D. Mills.** 2006. Comparative genomics of the lactic acid bacteria. Proc. Natl. Acad. Sci. USA **103:**15611–15616.
67. **Marchler-Bauer, A., J. B. Anderson, P. F. Cherukuri, C. DeWeese-Scott, L. Y. Geer, M. Gwadz, S. He, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, C. A. Liebert, C. Liu, F. Lu, G. H. Marchler, M. Mullokandov, B. A. Shoemaker, V. Simonyan, J. S. Song, P. A. Thiessen, R. A. Yamashita, J. J. Yin, and S. H. Bryant.** 2005. CDD: a Conserved Domain Database for protein classification. Nucleic Acids Res. **33:**D192–D196.
68. **McHardy, A. C., A. Goesmann, A. Puhler, and F. Meyer.** 2004. Development of joint application strategies for two microbial gene finders. Bioinformatics **20:**1622–1631.
69. **Merkl, R.** 2004. SIGI: score-based identification of genomic islands. BMC Bioinformatics **5:**22.
70. **Meyer, F., A. Goesmann, A. C. McHardy, D. Bartels, T. Bekel, J. Clausen, J. Kalinowski, B. Linke, O. Rupp, R. Giegerich, and A. Puhler.** 2003. GenDB—an open source genome annotation system for prokaryote genomes. Nucleic Acids Res. **31:**2187–2195.
71. **Mierau, I., and M. Kleerebezem.** 2005. 10 years of the nisin-controlled gene expression system (NICE) in *Lactococcus lactis*. Appl. Microbiol. Biotechnol. **68:**705–717.
72. **Mierau, I., P. Leij, I. van Swam, B. Blommestein, E. Floris, J. Mond, and E. Smid.** 2005. Industrial-scale production and purification of a heterologous protein in *Lactococcus lactis* using the nisin-controlled gene expression system NICE: the case of lysostaphin. Microb. Cell Factories **4:**15.
73. **Mulder, N. J., R. Apweiler, T. K. Attwood, A. Bairoch, D. Barrell, A. Bateman, D. Binns, M. Biswas, P. Bradley, P. Bork, P. Bucher, R. R. Copley, E. Courcelle, U. Das, R. Durbin, L. Falquet, W. Fleischmann, S. Griffiths-Jones, D. Haft, N. Harte, N. Hulo, D. Kahn, A. Kanapin, M. Krestyaninova, R. Lopez, I. Letunic, D. Lonsdale, V. Silventoinen, S. E. Orchard, M. Pagni, D. Peyruc, C. P. Ponting, J. D. Selengut, F. Servant, C. J. A. Sigrist, R. Vaughan, and E. M. Zdobnov.** 2003. The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res. **31:**315–318.
74. **Niven, C. F.** 1944. Nutrition of *Streptococcus lactis*. J. Bacteriol. **447:**343–350.
75. **O'Flaherty, S., A. Coffey, W. Meaney, G. F. Fitzgerald, and R. P. Ross.** 2005. The recombinant phage lysin LysK has a broad spectrum of lytic activity against clinically relevant staphylococci, including methicillin-resistant *Staphylococcus aureus*. J. Bacteriol. **187:**7161–7164.
76. **Reiter, B., and J. D. Oram.** 1962. Nutritional studies on cheese starters. I. Vitamin and amino acid requirements of single strain starters. J. Dairy Res. **29:**63–77.
77. **Rezaïki, L., B. Cesselin, Y. Yamamoto, K. Vido, E. van West, P. Gaudu, and A. Gruss.** 2004. Respiration metabolism reduces oxidative and acid stress to improve long-term survival of *Lactococcus lactis*. Mol. Microbiol. **53:**1331–1342.
78. **Rowland, S. J., and K. G. Dyke.** 1989. Characterisation of the staphylococcal beta-lactamase transposon Tn*552*. EMBO J. **8:**2761–2773.
79. **Russell, W. M., and T. R. Klaenhammer.** 2001. Efficient system for directed integration into the *Lactobacillus acidophilus* and *Lactobacillus gasseri* chromosomes via homologous recombination. Appl. Environ. Microbiol. **67:**4361–4364.
80. **Salzberg, S. L., A. L. Delcher, S. Kasif, and O. White.** 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. **26:**544–548.
81. **Sanz, Y., F. C. Lanfermeijer, M. Hellendoorn, J. Kok, W. N. Konings, and B. Poolman.** 2004. Two homologous oligopeptide binding protein genes (*oppA*) in *Lactococcus lactis* MG1363. Int. J. Food Microbiol. **97:**9–15. [Erratum, **102:**121, 2005.]
82. **Schleifer, K. H.** 1987. Recent changes in the taxonomy of lactic acid bacteria. FEMS Microbiol. Rev. **46:**201–203.
83. **Schleifer, K. H., J. Kraus, C. Dvorak, R. Kilpperblaz, M. D. Collins, and W.**

Fischer. 1985. Transfer of *Streptococcus lactis* and related streptococci to the genus *Lactococcus* gen. nov. Syst. Appl. Microbiol. **6:**183–195.
84. **Siezen, R., J. Boekhorst, L. Muscariello, D. Molenaar, B. Renckens, and M. Kleerebezem.** 2006. *Lactobacillus plantarum* gene clusters encoding putative cell-surface protein complexes for carbohydrate utilization are conserved in specific gram-positive bacteria. BMC Genomics **7:**126.
85. **Smid, E. J., D. Molenaar, J. Hugenholtz, W. M. de Vos, and B. Teusink.** 2005. Functional ingredient production: application of global metabolic models. Curr. Opin. Biotechnol. **16:**190–197.
86. **Staden, R., K. F. Beal, and J. K. Bonfield.** 2000. The Staden package, 1998. Methods Mol. Biol. **132:**115–130.
87. **Steen, A., G. Buist, K. J. Leenhouts, M. E. Khattabi, F. Grijpstra, A. L. Zomer, G. Venema, O. P. Kuipers, and J. Kok.** 2003. Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. J. Biol. Chem. **278:**23874–23881.
88. **Steidler, L., W. Hans, L. Schotte, S. Neirynck, F. Obermeier, W. Falk, W. Fiers, and E. Remaut.** 2000. Treatment of murine colitis by *Lactococcus lactis* secreting interleukin-10. Science **289:**1352–1355.
89. **Steidler, L., S. Neirynck, N. Huyghebaert, V. Snoeck, A. Vermeire, B. Goddeeris, E. Cox, J. P. Remon, and E. Remaut.** 2003. Biological containment of genetically modified *Lactococcus lactis* for intestinal delivery of human interleukin 10. Nat. Biotechnol. **21:**785–789.
90. **Steidler, L., and K. Vandenbroucke.** 2006. Genetically modified *Lactococcus lactis*: novel tools for drug delivery. Int. J. Dairy Technol. **59:**140–146.
91. **Suzek, B. E., M. D. Ermolaeva, M. Schreiber, and S. L. Salzberg.** 2001. A probabilistic method for identifying start codons in bacterial genomes. Bioinformatics **17:**1123–1130.
92. **Sybesma, W., C. Burgess, M. Starrenburg, D. van Sinderen, and J. Hugenholtz.** 2004. Multivitamin production in *Lactococcus lactis* using metabolic engineering. Metabol. Eng. **6:**109–115.
93. **Sybesma, W., M. Starrenburg, L. Tijsseling, M. H. N. Hoefnagel, and J. Hugenholtz.** 2003. Effects of cultivation conditions on folate production by lactic acid bacteria. Appl. Environ. Microbiol. **69:**4542–4548.
94. **Tatusov, R., N. Fedorova, J. Jackson, A. Jacobs, B. Kiryutin, E. Koonin, D. Krylov, R. Mazumder, S. Mekhedov, A. Nikolskaya, B. S. Rao, S. Smirnov, A. Sverdlov, S. Vasudevan, Y. Wolf, J. Yin, and D. Natale.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4:**41.
95. **Teuber, M.** 1995. The genus *Lactococcus*, p. 173–234. *In* B. J. B. Wood and W. H. Holzapfel (ed.), The genera of lactic acid bacteria. Blackie Academic and Professional, Glasgow, United Kingdom.
96. **Tynkkynen, S., G. Buist, E. Kunji, J. Kok, B. Poolman, G. Venema, and A. Haandrikman.** 1993. Genetic and biochemical characterization of the oligopeptide transport system of *Lactococcus lactis*. J. Bacteriol. **175:**7523–7532.
97. **van Hijum, S. A. F. T., A. L. Zomer, O. P. Kuipers, and J. Kok.** 2003. Projector: automatic contig mapping for gap closure purposes. Nucleic Acids Res. **31:**e144.
98. **van Hylckama Vlieg, J. E. T., J. L. W. Rademaker, H. Bachmann, D. Molenaar, W. J. Kelly, and R. J. Siezen.** 2006. Natural diversity and adaptive responses of *Lactococcus lactis*. Curr. Opin. Biotechnol. **17:**183–190.
99. **van Kranenburg, R., M. Kleerebezem, J. van Hylckama Vlieg, B. M. Ursing, J. Boekhorst, B. A. Smit, E. H. E. Ayad, G. Smit, and R. J. Siezen.** 2002. Flavour formation from amino acids by lactic acid bacteria: predictions from genome sequence analysis. Int. Dairy J. **12:**111–121.
100. **van Niel, E. W. J., and B. Hanhn-Hagerdal.** 1999. Nutrient requirements of lactococci in defined growth media. Appl. Microbiol. Biotechnol. **52:**617–627.
101. **Veith, B., C. Herzberg, S. Steckel, J. Feesche, K. H. Mauer, P. Ehrenreich, S. Baumer, A. Henne, H. Liesegang, R. Merkl, A. Ehrenreich, and G. Gottschalk.** 2004. The complete genome sequence of *Bacillus licheniformis* DSM13, an organism with great industrial potential. J. Mol. Microbiol. Biotechnol. **7:**204–211.
102. **Vido, K., D. le Bars, M.-Y. Mistou, P. Anglade, A. Gruss, and P. Gaudu.** 2004. Proteome analyses of heme-dependent respiration in *Lactococcus lactis*: involvement of the proteolytic system. J. Bacteriol. **186:**1648–1657.
103. **Zhou, C. E., J. Smith, M. Lam, A. Zemla, M. D. Dyer, and T. Slezak.** 2007. MvirDB—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res. **35:**D391–D394.